線形回帰モデルの母数と共変量への加法的共変量測定誤差の影響

放射線影響研究所（放影研）統計部：中島栄二

[はじめに] 線形回帰モデルの共変量測定誤差には、古典誤差（classical error）と、バークソン平均化誤差（Berkson averaging error）がある。この発表では、共変量誤差の回帰母数に対する影響を“真のモデル”と“実際のモデル”という二つの概念の下で考える。“真のモデル”とは未知で真の共変量を用いた仮想的なモデルであり、“実際のモデル”とは誤差を含む観測共変量を用いて理論的に導かれたモデルである。共変量誤差が無い回帰では、二つの回帰モデルは真のモデルに一致する。これらのモデルと、データ解析で用いられる母数節約的な“作業モデル”とは異なりうる。真のモデルと実際のモデルでは、真の共変量、共変量誤差、およびモデル誤差の間の独立性を仮定し、モデル誤差の等分散性は必ずしも仮定しない。このとき次の理論的結果を得る。

[理論的結果] 二つの場合（ケース１とケース２）を考える。真のモデルが、ある密度関数に従う真の連続共変量の上の単回帰モデルとして、実際のモデルの下での古典共変量誤差の母数に対する影響と共変量に対する影響を調べた。ケース１では、真の共変量密度が、ある区間上で基準化された二次指数分布族（対数密度が変数の高々二次式となる分布）、つまり、ある区間上で基準化された正規分布、指数分布、および一様分布の下で、古典共変量誤差がある場合、実際の回帰モデルは母数と共変量の両方について線形となる。この時、傾き母数は零の方向に縮小し、縮小係数は真の共変量分散と共変量誤差分散により決まる。

ケース２では、もっと一般に、どんな既知の共変量分布であっても、実際の回帰モデルは、観察された共変量の既知の非線形関数の上の線形回帰モデルとなる。この場合、その既知非線形関数は、真の共変量の分布と測定誤差（正規加法誤差の場合は誤差分散）によって決まる。従って、応用においては真の共変量分布の（必ずしもパラメトリックとは限らない）密度推定が重要な問題となる。独立なバークソン平均化誤差をもつ共変量誤差の下では、実際の線形回帰モデルは、共変量分布に関わりなく真の線形回帰モデルとなり母数も一致する。

[シミュレーション] ケース１の場合では、真の共変量が一様分布の場合で、古典共変量誤差により真の線形回帰が、傾き母数が縮小した実際の線形回帰となる場合のシミュレーションによる結果を示す。ケース２の場合は、共変量の分布が三角分布のような簡単な分布を用いて、真のモデルと実際のモデルが異なる例を理論計算により示す。

[可能性のある実例とコメント] 有限なサンプル数で、作業モデルを用いて、真のモデルと実際のモデルの違いを、決定論的に言うことは難しいが、計測誤差の影響の現れと考えられる実例（見かけ上低線量域での放射線高感受性を示す例）を引用し、これら二つのモデルが異なる実例がある可能性を示唆する予定である。